Created by Neil Yetz & Gemma Wallace

Module 9 Lab Activity: Essential Concepts in Multiple Regression
PSY 652 Research Methods
Oct 30, 2019

Description of the dataset:

A research team sought to examine factors associated with 21st birthday drinking among female students at a large University. Female students who were nearing age 21 and self-classified as regular drinkers were eligible for the study. In total, 200 students were recruited and agreed to take part in the study. Students were instructed to report to the lab two weeks prior to their 21st birthday. During this lab session, students completed a brief survey that measured alcohol use during the past month (using the Timeline Follow Back Method) and their weight was recorded. One week prior to their 21st birthday, participants were sent a link for an online survey to measure positive alcohol expectancies for drinking on their 21st birthday. Within three days prior to their 21st birthday, students reported to the lab and were given a diary-based data collection form to record several items on their 21st birthday. Students were instructed to record the food that they consumed during the day, the degree to which they were in a partying mood just prior to the celebration, and the quantity and type of drinks that they consumed during the first two hours of the celebration. The students were also given a small breathalyzer machine to measure BAC 2 hours after consumption of their first drink.
The dataset called bac_obs.csv contains the data:

- *weight: weight in kilograms*
- *alcexp: positive alcohol expectancy for drinking on the impending 21st birthday, a multi-item scale that ranges from 1-7, where a higher score indicates more positive expectations about the role alcohol will play*
- *typ_drks: the number of standard alcohol drinks consumed in the past 30 days*
- *pmood: a rating on a scale from 1-9 on the respondent's mood to party on the 21st birthday, where 1 means never been less in the mood to party, and 9 means never been more in the mood to party*
- *absorb: a score calculated from the food diaries to determine how full the participant was when they began drinking, the score rang- es from 1 to 8, where 1 means a completely full stomach, and 8 means a completely empty stomach*
- *alc_gm: a score calculated from the drinking diary to estimate the grams of alcohol consumed on the 21st birthday*
- *bac: the participant's blood alcohol content, measured as grams of alcohol per deciliter of blood on the 21st birthday*

1. Download the "bac_obs.csv" dataset from the Module 9 Lab dropbox folder and save it into a project folder.
2. Create a new R notebook from your project file and name it "bac_notebook"
3. Create a new R chunk with a first level header: "Load Libraries"
    a. load the psych, olsrr, and tidyverse packages, GGally (you may need to install GGally)

4. Create a new R chunk with a first level header: “Import Data”
    a. read in the “bac_obs.csv” dataset, assign it to an object named “bac”
5. Create a new R chunk with a first level header: “Describe data”
    a. Use any method to obtain descriptives for the bac dataset
6. Create a new R chunk with a first level header: “Mutate BAC variable”
    a. In this chunk create a new version of the bac variable called bac100 (bac100 = bac*100) and save this to the bac dataframe. We start here because bac is a very small number and multiplying it by a constant won’t change the overall model but will make our output easier to read.
    b. Confirm that this worked by opening the bac dataframe.
7. Create a new R chunk with the first level header: “Get Scatterplot Matrix”
    a. Copy and paste the following code into a new R chunk to create a scatterplot matrix of the variables bac100, typ_drks, weight, and alc_gm.

```
scatterplot2 <- ggpairs(bac, columns = c("bac100", "typ_drks", "weight", "alc_gm"), upper = list(continuous = wrap("cor", size=5)), title = "Bivariate Relationship of Key Variables", progress = ggmatrix_progress(clear = TRUE))

print(scatterplot2)
```

    (Hint: iniclude “show message = FALSE in the top of your R chunk to avoid jargon output).
    b. In the white space, write 1-2 sentences describing the scatterplot and correlation matrix and note anything that stands out.
8. Write a first level header: “Fit regression models”
    a. Create a second level header (“Regress bac100 on weight and typ_drks”) and a new code chunk. Regress bac100 on weight (i.e., run a model in which weight and typ_drks predict bac100) and ask for the output using the ols_regress function.
    b. Create a second level header (“Regress bac100 on alc_gm and typ_drks”) and a new code chunk. Regress bac100 on alc_gm and typ_drks and ask for the output using the ols_regress function.
    c. Create a second level header (“Regress bac100 on weight, alc_gm, and typ_drks”) and a new code chunk. Regress bac100 on weight, alc_gm, and typ_drks, and ask for the output using the old_regress function. Study the output that you have generated so far.
    d. For each of these three MLR models, write sentences in the white space to interpret the following: the intercept and the beta estimates, confidence intervals, and *p* values for all of the predictors.
    e. Examine the output for these three models and answer the following questions in the white spacee:
        i. How do the R^2 values vary across the three models? Why does the MLR with three predictors have the highest R^2?

   ii. How do the slopes, confidence intervals, and associated *p* values for each predictor vary across the three models? Why does the slope for each variable change when there are different combinations of predictors in the model?

9. Write a first level header: “Run regression through correlation matrix”
   a. Create a new R chunk with a second level header: “Subset data for correlation matrix”
      i. Create a subset of the bac dataframe that only includes the following variables: bac100, alc_gm, and weight. Name this new dataframe “bac_subset.”
      ii. In the same code chunk, create a correlation matrix of the bac_subset dataframe and save it to an object called “cor_matrix.” View the correlation matrix to ensure that this worked.
   b. Create a new R chunk with a second level header: “Calculate regression from correlation matrix”
      i. Use the set.cor function to regress bac100 on weight and alc_gm.
      Hint: set.cor(y = ("*outcome_variable*"), x = c("*predictor1*”, "*predictor2*"), z = NULL, data = *dataframe*)
      ii. Examine the slope for each predictor variable (in the 2nd window) and write a short sentence that compares these values to the estimates from the MLR you conducted above.
      iii. Examine the R and R2^2 values (in the 3rd window) and write a short sentence that compares these values to the output from the MLR you conducted above.
      Note: Don’t interpret the Ruw and R2uw values in this window (these are the *unweighted* values the regression coefficients)
      iv. In the white space, write 1-2 sentences to answer the following question: Why might you prefer to use the correlation matrix method to compute regression estimates over the MLR method?